Kongeriget Danmark

CERTIFIED COPY OF PRIORITY DOCUMENT

Patent application No.: PA 2000 01259

Date of filing: 25 August 2000

Applicant: Novo Nordisk A/S
Novo Allé
DK-2880 Bagsværd

Schering AG
D-13342 Berlin

This is to certify the correctness of the following information:

The attached photocopy is a true copy of the following information:

- The specification, claims, SEQ ID, abstract and sequence listing as filed with the application on the filing date indicated above.

Patent- og Varemærkestyrelsen
Kongeriget Danmark

Patent- og
Varemærkestyrelsen
Erhvervsministeriet

TAASTRUP 20 August 2001

Karin Schlichting
Head Clerk

BACKGROUND OF THE INVENTION

Receptors are defined as proteinaceous macromolecules that perform a signal transducing function upon ligand binding. Many receptors are located on the outer cell membrane, others are located intracellularly. The substance which is bound by the receptor is called a ligand, a term which is definitionally meaningful only in terms of its counterpart receptor. The term "ligand" does not imply any particular molecular size or other structural or compositional feature other than that the substance in question is capable of binding, cleaving or otherwise interacting with the receptor in such a way that the receptor conveys information about the presence of the ligand to a target molecule. Stated alternatively, not all substances capable of binding a receptor are ligands, but all ligands are capable of binding a receptor. Receptors do not include such substances as immunoglobulins.

Receptors are believed to function by a process variously termed activation or signal transduction. A ligand binds to the ligand binding domain in such a way that the conformation of the receptor molecule changes. This conformational change, called activation, modifies the effect of the receptor on cytoplasmic components. Among changes brought about by receptor activation are changes in or development of receptor enzymatic activity.

Signalling proteins such as cAMP, IP3, kinases, and phosphatases are proteins ubiquitously found in all tissues. These proteins cascade by various pathways, the stimulus from ligand/receptor interaction down stream to cellular events, typically changing the enzymatic activity or functional state of effector molecules.

The pharmaceutical industry in recent years has oriented its research to focus on the role of receptors in disease or injury and to design drugs, generally low molecular weight substances, that are capable of binding to the receptors. Drugs identified in this initial screen are then tested for the activity *in vivo* or in tissue explants. As a result, conventional techniques do not lend themselves to large-scale screening. Tissue samples or isolated cells containing the target receptors, for example ovarian tissue, are costly to obtain, present in limited quantity, and difficult to maintain in a functionally viable state. Additionally, it is often difficult to reliably and reproducibly administer the candidate drug to tissue samples. Screening assays using primary explants in tissue culture are undertaken in larger scale than is possible with

tissue samples. However, it is more difficult to assay physiological effect and the assays are subject to interference from many sources, for example culture media or cultivation conditions. Finally, assays using receptors isolated from natural materials have the disadvantage that the receptor is subject to natural variability and suitable natural sources may not always be available. It is an object herein to provide readily reproducible, simple assay systems that can be practiced on a large scale for determining not only ligand binding but also the character of the binding as agonistic or antagonistic.

Similarly, meaningful clinical diagnoisis often depends upon the assay of biologically active ligand without interference from inactive forms of the ligand, for example, ligands that have been subject to enzymatic or other processes in the test subject that change or even eliminate the activity of the ligand. Immunoassay methods are widely used in determining ligands in test samples. However, it is often quite difficult to identify antibodies that are able to discriminate between the active and inactive forms of a ligand. Receptors have frequently been used in place of antibodies as analyte binding reagents. However, not all substances that bind to receptors are necessarily capable of inducing receptor activity, i.e. active biologically. It is an object herein to provide a method that will identify ligands in clinical test samples which are active in inducing or inhibiting signal transduction by their receptors.

Cytoplasmic proteins can act as receptors or signalling molecules in cascading the stimulus from the ligand to cellular events. Various receptor or signalling protein types make use of different path ways (for example small G proteins, calcium fluxes, phospatases and lipases) all of them resulting in changes of enzymatic activity or gene transcription.

Meiotic activating sterols (MAS) constitute active signalling molecules first identified in follicular fluid and in bull testicular tissue. The sterols are described by Byskov 1995 and Grøndahl et al. (Biol. Reprod. 58 (1998), 1297 *et seq.*) and in WO 96/00235, 96/27658, 97/00884, 98/28323, 98/54965 and 98/55498, more specifically in Claim 1 thereof, as being potent activators of the meiotic process. No receptors or signalling proteins have been described to directly or indirectly signalling the meiotic effect of MAS sterols. Before this invention, the presence of the nature of a putative MAS receptor protein or a signalling protein has not been previously been identified, although its presence has been suggested, for example, by Grøndahl et al. (Biol. Reprod. 62 (2000), 775 *et seq.*).

There remains considerable need for isolated and purified MAS receptor or MAS signalling protein, as well as systems capable of expressing MAS receptor or MAS signalling protein separate from other receptors. Further, it would be desirable to specifically identify the presence of MAS receptor or MAS signalling protein in cells and tissues, thereby avoiding the time-consuming, complex and non-specific functional pharmacological assays. It would also be desirable to screen and develop new agonists and/or antagonists specific for MAS receptor or MAS signalling protein for the use of antiinfertility or contraception drugs, but to date this has not been possible. Quite surprisingly, the present invention fulfils these and other related needs.

SUMMARY OF THE INVENTION

Now, the present invention provides the nucleotide sequence of the receptors or signalling proteins of meiotic acting sterols. The present invention provides isolated and substantially pure MAS receptors or MAS signalling proteins and fragments thereof. These receptors or signalling proteins have been shown to be involved in the gamete maturation process induced by FF-MAS, specifically inducing, upon ligand activation, germinal vesicle breakdown (GVB) in mouse oocyte cultured in vitro.

A MAS receptor is any protein related to the protein SAM1a or SAM1b that possess the same functional characteristic regarding the interaction with 3β-hydroxy-4,4-dimethylcholest-8,14,24-triene (hereinafter designated FF-MAS) or other endogenous meiosis activating sterols, for example, 3β-hydroxycholest-8,14-diene; 3β-hydroxy-4,4-dimethylcholest-8,24-diene; and 3β-hydroxycholest-8,24-diene, or their metabolites (as ligand). Functional characteristics include binding, receptor activation, and subsequent germinal vesicles breakdown (GVB) in oocytes. The amino acid sequence of SAM1a and SAM1b is stated in SEQ ID NO: 2 and SEQ ID NO: 4, below.

Having provided such receptors or signalling proteins in isolated or purified form, the invention also provides antibodies to the receptor or signalling protein, in the form of antisera and/or monoclonal antibodies.

In another aspect, the invention provides the ability to produce the MAS receptor or MAS signalling protein and polypeptides or fragments thereof by recombinant means. The expressed receptor or signalling protein or fragments may or may not have the biological activity of native receptor or signalling protein. Accordingly, isolated and purified polynucleotides are described which code for the receptor or signalling protein and fragment thereof, where the polynucleotides may be in the form of DNA, such as cDNA, or RNA. Based on these sequences, probes may be used to hybridise and identify these and related genes which encode MAS receptors or MAS signalling proteins. The probes may be full length cDNA or as small as form 14 to 25 nucleotide, more often though from about 40 to about 50 or more nucleotides.

In related embodiments, the invention concerns DNA constructs which comprise a transcriptional promoter, a DNA sequence which encodes the receptor or signalling protein or fragment, and a transcriptional terminator, each operably linked for expression of the receptor or signalling protein. For expression, the construct may also contain at least one signal sequence. Further, for large scale production, the expressed receptor or signalling protein may also be isolated from the cells by, for example, immunoaffinity purification.

Cells or bacteria which express the MAS receptor or MAS signalling proteins may also be used to identify compounds which can alter the receptor or signalling protein-mediated metabolism of a cell. Compounds may be screened for binding to the receptor or signalling protein, and/or for effecting a change in receptor or signalling protein-mediated metabolism in the host cell. Agonists and/or antagonists of the MAS receptor or MAS signalling proteins may also be screened in cell-free systems using purified receptor or signalling proteins or binding fragments thereof for the effect on ligand/receptor interaction or ligand/signalling protein interaction, or using reconstituted systems such micelles which also provide the ability to assess metabolic changes.

In yet other embodiments, the invention relates to methods for diagnosis, where the presence of a mammalian MAS receptor or MAS signalling protein in a biological sample may be determined. For example, a monospecific antibody which specifically binds the receptor or signalling protein is incubated with the sample under conditions conducive to immune com-

plex formation, which complexes are then detected, typically by means of a label such as an enzyme, fluorophore, radionuclide, chemiluminiscer, particle, or a second labelled antibody. Thus, means are provided for immunohistochemical staining of tissues, including ovarian or testicular tissues, for the subject receptor or signalling proteins.

Based upon the similarity in sequence and the shared presence of a sterol binding domain at the protein level, the receptor proteins or signalling proteins of this invention can be said to belong to a novel super family of oxysterol binding proteins (hereinafter designated OSPB) recently published in *J.Lipid.Res.* 40 (1999), 2204. No function whatsoever in gamete maturation of either gender or regulation of any meiotic processes has been assigned to this OSPB family.

BRIEF DESCRIPTION OF THE FIGURES

SEQ ID NO: 1 and SEQ ID NO: 3 are the nucleotides of the cDNA from two MAS receptors or signalling peptides having the amino acid sequences stated in SEQ ID NO: 2 and SEQ ID NO: 4, respectively.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention presents the means to identify agonists, and antagonists of the MAS receptor/ligand interaction or MAS signalling protein/ligand interaction by providing isolated MAS receptor or MAS signalling protein. The term "MAS receptor" refers to any proteins either derived from a naturally occurring MAS receptor, or which shares significant structural and functional characteristics peculiar to a naturally occurring MAS receptor. Such a receptor may result when regions of a naturally occurring receptor are deleted or replaced in such a manner as to yield a protein having a similar function. Homologous sequences, allelic variations, and natural mutants; induced point, deletion, and insertion mutants; alternatively expressed variants; proteins encoded by DNA which hybridise under high or low stringency conditions to nucleic acids which encode naturally occurring MAS receptor; proteins retrieved from naturally occurring materials; and closely related proteins retrieved by antisera

directed against MAS receptor proteins are also included. Similarly, this applies to MAS signalling proteins.

By MAS receptor "ligand" is meant a molecule capable of being bound by the ligand-binding domain of MAS receptor, a MAS receptor analogue, or chimeric MAS receptor as generally described in US Pat. No. 4,859,609, incorporated by reference herein. The molecule may be chemically synthesised or may occur in nature. Ligands may be grouped into agonists and antagonists. Agonists are those molecules whose binding to a receptor induces the response pathway within a cell. Antagonists are those molecules whose binding to a receptor blocks the response pathway within a cell.

By "isolated" MAS receptor or MAS signalling protein is meant to refer to MAS receptor or MAS signalling protein which is in other than its native environment such as a mammalian oocyte, including, for example, substantially pure MAS receptor as defined herein below. More generally, isolated is meant to include MAS receptor or MAS signalling protein as a heterologous component of a cell or other system. For example, MAS receptor or MAS signalling protein may be expressed by a cell transfected with a DNA construct which encodes MAS receptor or MAS signalling protein, separated from the cell and added to micelles which contain other selected receptor or signalling proteins.

By purified MAS receptor or MAS signalling protein is meant MAS receptor or MAS signalling protein having a purity of at least 50%, preferably at least 80%, more preferred at least 90% (w/w).

In another aspect, the invention provides means for regulating the MAS receptor/ligand interaction or MAS signalling protein/ligand interaction, and thus treating, therapeutically and/or prophylactically, a disorder which can be linked directly or indirectly to MAS receptor or MAS signalling protein or to its ligands, such as FF-MAS. By virtue of having the receptor or signalling protein of the invention, agonists or antagonists may be identified which stimulate or inhibit the interaction of ligand with MAS receptor or with MAS signalling protein. With either agonists or antagonists, the metabolism and reactivity of cells which express the receptor or signalling protein are controlled, thereby providing a means to control meiosis in order to treat infertility or to achieve a novel principle of contraception.

Thus, the invention provides screening procedures for identifying agonists or antagonists of events mediated by the ligand/MAS receptor or ligand/MAS signalling protein interaction. Such screening assays may employ a wide variety of formats, depending to some extent on which aspect of the ligand, receptor or signalling protein interaction is targeted. For example, such assays may be designed to identify compounds which bind to the receptor or signalling protein and thereby block or inhibit interaction of the receptor or signalling protein and thereby block or inhibit interaction of the receptor or signalling protein with the ligand. Other assays can be designed to identify compounds which can substitute for ligand and therefore stimulate MAS receptor-mediated or MAS signalling protein-mediated intracellular pathways. Yet other assays can be used to identify compounds which inhibit or facilitate the association of MAS receptor or MAS signalling protein to FF-MAS and thereby mediate the cellular response to MAS receptor or MAS signalling protein ligand.

In one functional screening assay, the initiation of fertilisation activation events are monitored in eggs which have been injected with, for example, mRNA which codes for MAS receptor or MAS signalling protein and subsequently exposed to selected compounds which are being screened, in conjunction with or apart form an appropriate ligand. See generally, Kline *et al.*, *Science* 241 (1988), 464-467, incorporated herein by reference.

The screening procedure can be used to identify reagents such as antibodies which specifically bind to the receptor or signalling protein and substantially affect its interaction with ligand, for example. The antibodies may be monoclonal or polyclonal, in the form of antiserum or monospecific antibodies, such as purified antiserum or monoclonal antibodies or mixtures thereof. For administration to humans, for example, as a component of a composition for in vivo diagnosis or imaging, the antibodies are preferably substantially human to minimise immunogenicity and are in substantially pure form. By substantially human is meant generally containing at least about 70% human antibody sequence, preferably at least about 80% human, and most preferably at least about 90-95% or more of a human antibody sequence to minimise immunogenicity in humans.

Antibodies which bind MAS receptor or MAS signalling protein may be produced by a variety of means. The production of non-human antisera or monoclonal antibodies, for example,

murine, lagomorpha equine, etc. is well known and may be accomplished by, for example, immunising the animal with the receptor or signalling protein molecule or a preparation containing a desired portion of the receptor or signalling protein molecule, such as that domain or domains which contributes to ligand binding. For the production of monoclonal antibodies, antibody-producing cells obtained from immunised animals are immortalised and screened, or screened first for the production of antibody which binds to the receptor or signalling protein and then immortalised. As the generation of human monoclonal antibodies to human MAS receptor or MAS signalling protein antigen may be difficult with conventional techniques, it may be desirable to transfer antigen binding regions of the non-human antibodies, for example, the $F(ab')_2$ or hypervariable regions, to human constant regions (Fc) or framework regions by recombinant DNA techniques to produce substantially human molecules. Such methods are generally known in the art and are described in, for example, US Pat. No. 4,816,397, and EP publications 173,494 and 239,400, which are incorporated herein by reference. Alternatively, one may isolate DNA sequences which code for a human monoclonal antibody or portions thereof that specifically bind to the human receptor or signalling protein by screening a DNA library from human B cells according to the general protocol outlined by Huse et al., *Science* 246 (1989), 1275-1281, incorporated herein by reference, and then cloning and amplifying the sequences which encode the antibody (or binding fragment) of the desired specificity.

In other embodiments, the invention provides screening assays conducted in vitro with cells which express the receptor or signalling protein. For example, the DNA which encodes the receptor or signalling protein or selected portions thereof may be transfected into an established cell line, for example, a mammalian cell line such as BHK and CHO, using procedures known in the art (see, for example, Sambrook et al, *Molecular Cloning, A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989, which is incorporated herein by reference). The receptor or signalling protein is then expressed by the cultured cells, and selected agents are screened for the desired effect on the cell, separately or in conjunction with an appropriate ligand such as FF-MAS or other MAS compounds. Means for amplifying nucleic acid sequences which may be employed to amplify sequences encoding the receptor or signalling protein or portions thereof are described in US Pat. Nos. 4,683,195 and 4,683,202, incorporated herein by reference.

In yet another aspect, the screening assays provided by the invention relate to transgenic mammals whose germ cells and somatic cells contain a nucleotide sequence encoding MAS receptor protein or signalling protein or a selected portion of the receptor which, for example, binds ligand. In yet a further aspect, the screening assays provided by the invention relate to transgenic mammals where the nucleotide sequence encoding a MAS receptor or a MAS signalling protein is molecularly targeted to produce knock out animals with the phenotypical loss of the specific MAS signalling function. Preferentially, the molecular knock out is tissue specific to gonadal tissue (ovary or testes) and is timely controlled in the development, thus inducible. There are several means by which a sequence encoding, for example, the human MAS receptor may be introduced into a non-human mammalian embryo or, alternatively, knocked out, some of which are described in, for example, US Pat. No. 4,736,866, Jaenisch, *Science* 240: 1468-1474 (1988) and Westphal et al., *Annu. Rev. Cell Biol.* 5: 181-196 (1989), which are incorporated herein by reference. The animal's cells then express the receptor or signalling protein and thus may be used as a convenient model for testing or screening selected agonists or antagonists.

In another aspect the invention concerns diagnostic methods and compositions. By means of having the MAS receptor or MAS signalling protein molecule and antibodies thereto, a variety of diagnostic assays are provided. For example, with antibodies, including monoclonal antibodies, to MAS receptor or MAS signalling protein, the presence and/or concentration of receptor or signalling protein in selected cells or tissues in an individual or culture of interest may be determined. These assays can be used in the diagnosis and/or treatment of diseases such as, for example, male infertility, female infertility, or by means of contraception in both gender.

Numerous types of immunoassays are available and are known to those skilled in the art, for example, competitive assays, sandwich assays, and the like, as generally described in, for example US Pat. Nos. 4,642,285; 4,376,110; 4,016,043; 3,879,262; 3,852,157; 3,850,752; 3,839,153; 3,791,932; and Harlow and Lane, *Antibodies, A Laboratory Manual*, Cold Spring Harbor Publications, N.Y. (1988), each incorporated by reference herein. In one assay format, MAS receptor or MAS signalling protein is identified and/or quantified by using labelled antibodies, preferably monoclonal antibodies which are reacted with brain tissues, for example, ovarian or testicular tissue, oocyte preparations, or semen samples, and determining the

specific binding thereto, the assay typically being performed under conditions conducive to immune complex formation. Unlabeled primary antibody can be used in combination with labels that are reactive with primary antibody to detect the receptor or signalling protein. For example, the primary antibody may be detected indirectly by a labelled secondary antibody made to specifically detect the primary antibody. Alternatively, the anti-MAS receptor-antibody or MAS signalling protein-antibody can be directly labelled. A wide variety of labels may be employed, such as radionuclides, particles (for example, gold, ferritin, magnetic particles, red blood cells), flourophores, chemiluminescers, enzymes, enzyme substrates, enzyme cofactors, enzyme inhibitors, and ligands (particularly haptens).

The RNA encoding the MAS receptor DNA or MAS signalling protein DNA may be directly detected in cells with a labelled synthetic oligonucleotide probe targeting the MAS receptor (RNA) or MAS signalling protein RNA in a hybridisation procedure. Also, the polymerase chain reaction (Saiki *et al.*, *Science* 239 (1988), 487, and US Pat. No. 4,683,195, each reference is hereby incorporated by reference) may be used to amplify DNA sequences, which are subsequently detected by their characteristic size on agarose gels, Southern blot of these gels using MAS receptor DNA or MAS signalling protein DNA or a oligonucleotide probe, or a dot blot using similar probes. The probes may comprise from about 14 nucleotides to about 25 or more nucleotides, preferably, 40 to 60 nucleotides, and in some instances a substantial portion or even the entire cDNA of MAS receptor or MAS signalling protein may be used. The probes are labelled with detectable signal, such as an enzyme, biotin, a radionuclide, fluorophore, chemiluminescer, and paramagnetic particle. High stringency in connection with hybridisation is obtained using the proper temperature and salt concentration.

Kits can also be supplied for use with the receptor or signalling protein of the subject invention in the detection of the presence of the receptor or signalling protein or antibodies thereto, as might be desired in the case of autoimmune disease. Thus, antibodies to MAS receptor or MAS signalling protein, preferably monospecific antibodies such as monoclonal antibodies, or compositions of the receptor or signalling protein may be provided, usually in lyophilised form in a container, either segregated or in conjunction with additional reagents, such as anti-antibodies, labels, gene probes, polymerase chain reaction primers and polymerase, and the like.

The following example is offered by way of illustration, not by limitation.

EXAMPLE 1

Microinjection of phosphorothionate oligonucleotides into mouse oocytes

Two antisense oligonucleotides (20 nucleotides) were utilized for microinjection: 5'-TCCACGATGGACGCCATCTT-3' and 5'-GCCAGCAGGAGAGCCATTCG-3', complementary to the kozak sequence of the mRNA encoded by the cDNA sequence herein designated *SAM1a* and *SAM1b*, respectively, both of which are defined in SEQ ID NO: 1 and SEQ ID NO: 3, respectively, shown below. In control experiments, the corresponding sense oligonucleotides were microinjected: 5'-AAGATGGCGTCCATCGTGGA-3' and 5'-CGAATGGCTCTCCTGCTGGC-3' for mRNA *SAM1a* and *SAM1b*, respectively. SAM1a antisense was co-injected with SAM1b antisense from a stock solution containing 1.25 µg/µl of each nucleotide in 10 % human serum albumin (hereinafter designated HSA) plus 5 mM Tris (pH value: 7.5). SAM1a sense was co-injected with SAM1b sense from a stock solution containing 1.25 µg/µl of each nucleotide in 10 % HSA plus 5 mM Tris (pH value: 7.5). Approximately 12 pg of each oligonucleotide (10 pl) were injected into the cytoplasma of individual germinal vesicle (GV)-stage oocytes loaded in a droplet of alpha-MEM supplemented with 0.8% HSA and 3 mM hypoxantine under mineral oil in a 35 mm petri dish on the stage of an inverted microscope. The oocytes were obtained from the ovaries of 21-24 days old mice following 48 hours priming with follicle stimulating hormone (hereinafter designated FSH) as described by Grøndahl *et al.* 1998 in *Biol. Reprod.* 58 (1998), 1297 *et seq.* Oocytes were sucked on to a holding pipette (120 µM outer diameter and 20 µm inner diameter) and an injection pipette (Eppendorph, Hamburg, Germany) was fitted to a pressure microinjector (Eppendorph, Hamburg, Germany). The pipette holder was attached to a piezoelectric positioning system (Burleigh, NY, USA) mounted on a motorized micromanipulator (Luigs and Neumann, Ratingen, Germany). The injection pipette was pushed against the zona pellucida, and then a piezoelectric pulse was given, moving the injection pipette 20 µm forward. During this movement the pipette penetrated the zona pelludica and oolema and then a brief pressure pulse was applied to release a volume of approximately 10 µl into the oocytes cy-

toplasm. Injected oocytes were placed in a CO_2 incubator at 37 C for 20 hours before resumption of meiosis was triggered by addition of 10 µM FF-MAS to the hypoxanthine containing medium. The effect of FF-MAS was evaluated after 24 hours of further incubation as the number of oocytes in germinal vesicle breakdown (GVBD). The rationale for the 20 hours cultivation period following injection of antisense oligonucleotides is to allow for degradation of mRNA coding for SAM1a and SAM1b protein. Consequently, when the level of MAS receptor protein or MAS signalling protein is reduced in the oocytes, the MAS response is blunted (from 100% to 50%, *vide* the table below).

Table 1

Oligonucleotide	GVBD/GVBD+GV 10 µM FF-MAS (24 hours)
SAM1a + SAM1b Antisense	13/25
SAM1a + SAM1b Sense	10/10
Non-injected oocytes	26/29

As shown in Table 1, GVBD was inhibited by 50% in antisense injected oocytes compared to control (i.e., sense injected and non-injected oocytes). This result indicates a selective inhibition of the mRNAs coding for SAM1a and SAM1b by the antisense probe. Furthermore, these results indicate that SAM1a and SAM1b proteins are crucial involved in the MAS signalling, since a functional knock out of *de novo* protein synthesis of these molecules partly disrupt the MAS signals in oocytes.

SAM1a and SAM1b are two closely related proteins originating from the same gene which possesses complementary functions regarding MAS signalling in oocytes.

What is claimed is:

1. An isolated and/or purified polynucleotide molecule which hybridises at high stringency to an oligonucleotide of 25 or more contiguous nucleotides of SEQ ID NO: 1 or SEQ ID NO: 3 and which polynucleotide codes for a) a MAS receptor or MAS signalling protein; or b) a ligand binding domain of a MAS receptor or MAS signalling protein.

2. The polynucleotide of the preceding claim, which is a RNA antisense sequence.

3. The polynucleotide of claim 1, which is a cDNA sequence.

4. The polynucleotide according to any one of the preceding claims, which encodes a polypeptide displaying MAS receptor or MAS signalling protein activity.

5. The polynucleotide of claim 1, which encodes a MAS receptor or MAS signalling protein having the amino acid sequence of SEQ ID NO: 2 or SEQ ID NO: 4.

6. A probe which comprises an oligonucleotide of 25 or more contiguous nucleotides of SEQ ID NO: 1 or SEQ ID NO: 3 capable of specifically hybridising with a gene which encodes a MAS receptor or MAS signalling protein, or allelic and species variants thereof.

7. The probe of the preceding claim, which comprises from about 40 to about 60 nucleotides in length.

8. The probe according to any one of the two preceding claims, which is labelled to provide a detectable signal.

9. A probe of at least 12 nucleotides, said probe being capable of hybridising with nucleic acids which encode a MAS receptor or MAS signalling protein.

10. A DNA construct comprising a DNA sequence which hybridises at high stringency to

an oligonucleotide of 25 or more contiguous nucleotides of SEQ ID NO: 1 or SEQ ID NO: 3 and which encodes a) a MAS receptor or MAS signalling protein; or b) a ligand binding domain of a MAS receptor or MAS signalling protein.

11. The DNA construct of the preceding claim, wherein the DNA sequence encodes a MAS receptor or MAS signalling protein having the amino acid sequence of SEQ ID NO: 2 or SEQ ID NO: 4.

12. A cultured cell line, yeast or bacteria transformed or transfected with a DNA construct which comprises a DNA sequence which hybridises at high stringency to an oligonucleotide of 25 or more contiguous nucleotides of SEQ ID NO: 1 or SEQ ID NO: 3 and which encodes a) a MAS receptor or MAS signalling protein; or b) a ligand binding domain or a transmembrane domain of a MAS receptor or MAS signalling protein.

13. The cell line, yeast or bacteria according to the preceding claim, which does not express endogenous MAS receptor or MAS signalling proteins.

14. An MAS receptor or MAS signalling protein, a peptide fragment thereof or a salt thereof, which is isolated and/or purified.

15. The isolated and/or purified protein of the preceding claim, comprising the amino acid sequence of SEQ ID NO: 2 or SEQ ID NO: 4.

16. An isolated antibody which specifically binds to a MAS receptor or MAS signalling protein.

17. The isolated antibody of the preceding claim wherein said antibody is a monoclonal antibody.

18. The isolated antibody according to any one of the preceding two claims, which blocks the binding of MAS to a MAS receptor or MAS signalling protein.

19. A hybridoma which produces a monoclonal antibody according to any of the two preceding claims.

20. A method for detecting the presence of a compound or a salt thereof which has affinity for a MAS receptor or MAS signalling protein, comprising the steps of a) contacting the compound with the MAS receptor or MAS signalling protein, a peptide fragment thereof or a salt thereof; and b) measuring the affinity of said compound for the MAS receptor or MAS signalling protein.

21. The method of the preceding claim for detecting the presence of MAS antagonists, comprising the steps of a) exposing a compound in the presence of a MAS agonist including MAS to a MAS receptor or MAS signalling protein coupled to a response pathway under conditions and for a time sufficient to allow binding of the compound to the receptor or MAS signalling protein and an associated response through the pathway; and b) detecting a reduction in the stimulation of the response pathway resulting from the binding of the compound to the MAS receptor or MAS signalling protein, relative to the stimulation of the response pathway by the MAS agonist alone and therefrom determining the presence of a MAS antagonist.

22. The method of any one of the two preceding claims for detecting the presence of MAS agonists, comprising the steps of a) exposing a compound in the presence of a MAS antagonist to a MAS receptor or MAS signalling protein coupled to a response pathway under conditions and for a time sufficient to allow binding of the compound to the receptor or MAS signalling protein and an associated response through the pathway; and b) detecting an increase of the stimulation of the response pathway resulting from the binding of the compound to the MAS receptor or MAS signalling protein, relative to the stimulation of the response pathway by the MAS antagonist alone and therefrom determining the presence of a MAS agonist.

23. A compound or a salt thereof which has affinity for the MAS receptor or a MAS signalling peptide and which compound or salt is detected by the method according to claim 20 or 21.

24. A method for producing a MAS receptor or MAS signalling protein having the amino acid sequence of SEQ ID NO: 2 or SEQ ID NO: 4, which comprises a) growing cells, yeast or bacteria transformed or transfected with a DNA construct which comprises a DNA sequence of SEQ ID NO: 1 or SEQ ID NO: 3 coding for the expression of the MAS receptor or MAS signalling protein, and b) isolating the receptor or MAS signalling protein from the cells.

25. The method according to the preceding claim, wherein the MAS receptor or MAS signalling protein is isolated by immunoaffinity purification.

26. A kit for screening a compound or a salt thereof which has affinity for a MAS receptor or MAS signalling protein, which contains the MAS receptor or MAS signalling protein, the peptide fragment thereof or a salt thereof.

SEQ ID NO: 1

SAM1a DNA

TTTGTTGTCATTGGCGGCTCCCAAGATGGCGTCCATCGTGGAAGGGCCGCTGAGCAAA
TGGACTAACGTGATGAAGGGATGGCAGTATCGTTGGTTCGTGCTGGACTACAATGCAG
GGCTGCTCTCCTACTACACGTCCAAGGACAAAATGATGAGAGGCTCTCGAAGAGGATG
CGTTAGACTCAGAGGAGCTGTGATTGGTATAGACGACGAGGACGACAGCACCTTCACA
ATCACTGTCGATCAGAAAACCTTCCACTTCCAGGCTCGAGATGCAGACGAGCGAGAGA
AGTGGATCCATGCCTTAGAAGAAACTATTCTTCGCCATACTCTTCAGCTTCAAGGTTTGG
ATTCAGGATTCATCCCCAGTGTCCAAGACTTTGATAAGAAACTTACCGAGGCTGACGCG
TACCTGCAGATCTTGATAGAACAATTAAAGCTTTTTGATGACAAGCTTCAAAATTGTAAA
GATGATGAACAGAGAAAGAAAGTTGAAACTCTCAAAGACACAACAAATAGCATGGTAGA
ATCAATTAAACACTGCATTGTGTTGCTACAGATTGCTAAAAGTACTATTAATCCTGTAGAT
GCAATATACCAGCCTAGTCCCTTGGAACCTGTGATCAGCACAATGCCTTCCCAGACTGC
CTTACCTCCAGAACCCGCTCAGTTGTGTAAGTCAGAGCAGCGTCCATCTTCCTTACCTG
TTGGACCTGTGTTAGCTACCTTGGGACATCATCAGACTCCAACACCAAATAGTACAGGC
AGTGGGAACTCACCACCTAGCAGCAGTCTGACTCCTCCCAGCCATGTCAACTTGTCTCC
AAATACAGTCCCAGAGTTCTCTTACTCTAGCAGTGAAGATGAGTTCTATGATGCTGATGA
ATTCCATCAAAGTGGCTCGTCCCCAAAGCGCTTAATAGATTCTTCTGGATCTGCCTCAGT
CTTGACACACAGCAGCTCCGGAAATAGCTTAAAACGCCCAGATACCACAGAGTCTCTGA
ATTCCTCCATGTCCAATGGCACAAGCGATGCTGATCTTTTTGACTCACATGACGACAGA
GATGATGATGGGGAGGCTGGGTCAGTGGAGGAGCACAAGAGCGTTATCATGCACCTCT
TATCACAAGTCAGGCTGGGGATGGACCTCACAAAGGTAGTTCTTCCAACGTTTATTCTC
GAGAGAAGATCTCTGTTAGAAATGTATGCAGACTTTTTCGCACATCCAGACCTGTTCGT
GAGCATTAGTGATCAGAAGGATCCCAGGGATCGAATGGTTCAGGTTGTGAAATGGTAC
CTCTCGGCCTTCCATGCAGGAAGGAGAGGATCGGTGGCCAAAAAGCCGTACAATCCTA
TTTTGGGTGAGATCTTTCAGTGTCACTGGACGTTGCCGAATGATACTGAAGAGAACGCA
GAGCTCGTTTCAGAAGGGCCGGTTCCCTGGGTTTCTAAGAACAGTGTAACATTTGTGGC
TGAGCAAGTTTCCCACCATCCGCCCATTTCAGCCTTTTATGCTGAGTGTTTTAACAAGAA
GATACAATTCAATGCTCATATCTGGACTAAATCAAAATTCCTTGGGATGTCAATTGGGGT
ACACAACATAGGTCAGGGCTGTGTCTCGTGTCTGGAGTACGATGAGCACTACATCCTCA
CGTTCCCCAATGGCTATGGAAGGTCTATCCTGACAGTGCCCTGGGTGGAATTGGGAGG

AGAATGCAATATCAACTGCTCCAAAACGGGTTACAGCGCAAACATCGTCTTCCACACTA
AGCCTTTCTATGGGGGCAAGAAGCACAGAATTACTGCAGAGATTTTTTCTCCGAATGAC
AAGAAATCCTTCTGCTCAATTGAAGGGGAATGGAATGGTATCATGTATGCAAAATACGC
AACAGGGGAAAACACTGTCTTTGTAGACACCAAGAAATTGCCTATAATCAAGAAAAAGG
TGAGGAAGTTGGAAGATCAGAATGAGTATGAGTCCCGCAGCCTTTGGAAGGATGTCAC
TTTCAATTTAAAAATCAGAGACATTGATGCAGCAACGGAAGCAAAGCACAGACTTGAAG
AAAGACAAAGAGCAGAAGCCCGAGAAAGGAAGGAGAAGGAAATTCAGTGGGAGACGA
GGCTCTTTCACGAGGATGGCGAATGCTGGGTTTACGATGAGCCTTTACTGAAGCGTCTT
GGTGCTGTGAAGCATTAGGCCGACAACCGAGTCCACACCTGGTGACCAGGGCAGTAG
GCGTAATTAAGCAACAATCGATCTTCCTTCAGGAGAGCTTGTCACTTCCTTCTTAACGCA
GTGGTTCCTATCTCAGGGATACTGGACTTGACGACACAGATGAACAATTAAAGTGGAAA
CCGCTTCCCTTTTCTCCTCTGTGGCAGTTACAATTTTGACTTCAGGCCTGAGAAAAACTT
CAGGTTTCGAAACATGACATCTCTTCCTTTTCCAGATCCCATGCTTTGAAAAATATTTATA
GACAGTTCCAGGTCTCAGCTTCCTGTCCTCTAGTTCTGCTGTTCGGGCATAAAATCTTTA
TCTCCAGTTCATATAATCTTGAGTTTTAGATAKACACACATGCGTAACAGCTGACAGTTTT
TCACAAGTACACCCACCTGTAAATACTGTAKCCTCAGTTTAGAAAATTAGTGCATGTATG
AAAATCAAGTGTTAGGAAATTTCATGGTTTCACCTATAACCTTTATTTTAGAATTGAACTA
TGATTAGATTGTATCTAAACCTGAAGTATAATTATATGCAGTGCTTCTTAAGGCTTCATAA
AATAATTTTCCAACCTTTTTAAAAAAAAAAAAAAA

SEQ ID NO: 2

SAM1a peptide

MASIVEGPLSKWTNVMKGWQYRWFVLDYNAGLLSYYTSKDKMMRGSRRGCVRLRGAVIGI
DDEDDSTFTITVDQKTFHFQARDADEREKWIHALEETILRHTLQLQGLDSGFIPSVQDFDKKL
TEADAYLQILIEQLKLFDDKLQNCKDDEQRKKVETLKDTTNSMVESIKHCIVLLQIAKSTINPV
DAIYQPSPLEPVISTMPSQTALPPEPAQLCKSEQRPSSLPVGPVLATLGHHQTPTPNSTGSG
NSPPSSSLTPPSHVNLSPNTVPEFSYSSSEDEFYDADEFHQSGSSPKRLIDSSGSASVLTHS
SSGNSLKRPDTTESLNSSMSNGTSDADLFDSHDDRDDDGEAGSVEEHKSVIMHLLSQVRL
GMDLTKVVLPTFILERRSLLEMYADFFAHPDLFVSISDQKDPRDRMVQVVKWYLSAFHAGR
RGSVAKKPYNPILGEIFQCHWTLPNDTEENAELVSEGPVPWVSKNSVTFVAEQVSHHPPISA
FYAECFNKKIQFNAHIWTKSKFLGMSIGVHNIGQGCVSCLEYDEHYILTFPNGYGRSILTVPW
VELGGECNINCSKTGYSANIVFHTKPFYGGKKHRITAEIFSPNDKKSFCSIEGEWNGIMYAKY
ATGENTVFVDTKKLPIIKKKVRKLEDQNEYESRSLWKDVTFNLKIRDIDAATEAKHRLEERQR
AEARERKEKEIQWETRLFHEDGECWVYDEPLLKRLGAVKH

SEQ ID NO: 3

SAM1b DNA

GGGCCCCTCCCCTGAGACGGCGGCCGGCCGCCGCGAATGGCTCTCCTGCTGGCCGC
TTGCGGAGGTTTGGATTCAGGATTCATCCCCAGTGTCCAAGACTTTGATAAGAAACTTA
CCGAGGCTGACGCGTACCTGCAGATCTTGATAGAACAATTAAAGCTTTTTGATGACAAG
CTTCAAAATTGTAAAGATGATGAACAGAGAAAGAAAGTTGAAACTCTCAAAGACACAACA
AATAGCATGGTAGAATCAATTAAACACTGCATTGTGTTGCTACAGATTGCTAAAAGTACT
ATTAATCCTGTAGATGCAATATACCAGCCTAGTCCCTTGGAACCTGTGATCAGCACAATG
CCTTCCCAGACTGCCTTACCTCCAGAACCCGCTCAGTTGTGTAAGTCAGAGCAGCGTCC
ATCTTCCTTACCTGTTGGACCTGTGTTAGCTACCTTGGGACATCATCAGACTCCAACACC
AAATAGTACAGGCAGTGGGAACTCACCACCTAGCAGCAGTCTGACTCCTCCCAGCCAT
GTCAACTTGTCTCCAAATACAGTCCCAGAGTTCTCTTACTCTAGCAGTGAAGATGAGTTC
TATGATGCTGATGAATTCCATCAAAGTGGCTCGTCCCCAAAGCGCTTAATAGATTCTTCT
GGATCTGCCTCAGTCTTGACACACAGCAGCTCCGGAAATAGCTTAAAACGCCCAGATAC
CACAGAGTCTCTGAATTCCTCCATGTCCAATGGCACAAGCGATGCTGATCTTTTTGACTC
ACATGACGACAGAGATGATGATGGGGAGGCTGGGTCAGTGGAGGAGCACAAGAGCGT
TATCATGCACCTCTTATCACAAGTCAGGCTGGGGATGGACCTCACAAAGGTAGTTCTTC
CAACGTTTATTCTCGAGAGAAGATCTCTGTTAGAAATGTATGCAGACTTTTTCGCACATC
CAGACCTGTTCGTGAGCATTAGTGATCAGAAGGATCCCAGGGATCGAATGGTTCAGGTT
GTGAAATGGTACCTCTCGGCCTTCCATGCAGGAAGGAGAGGATCGGTGGCCAAAAAGC
CGTACAATCCTATTTTGGGTGAGATCTTTCAGTGTCACTGGACGTTGCCGAATGATACT
GAAGAGAACGCAGAGCTCGTTTCAGAAGGGCCGGTTCCCTGGGTTTCTAAGAACAGTG
TAACATTTGTGGCTGAGCAAGTTTCCCACCATCCGCCCATTTCAGCCTTTTATGCTGAGT
GTTTTAACAAGAAGATACAATTCAATGCTCATATCTGGACTAAATCAAAATTCCTTGGGAT
GTCAATTGGGGTACACAACATAGGTCAGGGCTGTGTCTCGTGTCTGGAGTACGATGAG
CACTACATCCTCACGTTCCCCAATGGCTATGGAAGGTCTATCCTGACAGTGCCCTGGGT
GGAATTGGGAGGAGAATGCAATATCAACTGCTCCAAAACGGGTTACAGCGCAAACATC
GTCTTCCACACTAAGCCTTTCTATGGGGGCAAGAAGCACAGAATTACTGCAGAGATTTT
TTCTCCGAATGACAAGAAATCCTTCTGCTCAATTGAAGGGGAATGGAATGGTATCATGTA
TGCAAAATACGCAACAGGGGAAAACACTGTCTTTGTAGACACCAAGAAATTGCCTATAAT
CAAGAAAAAGGTGAGGAAGTTGGAAGATCAGAATGAGTATGAGTCCCGCAGCCTTTGG

AAGGATGTCACTTTCAATTTAAAAATCAGAGACATTGATGCAGCAACGGAAGCAAAGCA
CAGACTTGAAGAAAGACAAAGAGCAGAAGCCCGAGAAAGGAAGGAGAAGGAAATTCAG
TGGGAGACGAGGCTCTTTCACGAGGATGGCGAATGCTGGGTTTACGATGAGCCTTTAC
TGAAGCGTCTTGGTGCTGTGAAGCATTAGGCCGACAACCGAGTCCACACCTGGTGACC
AGGGCAGTAGGCGTAATTAAGCAACAATCGATCTTCCTTCAGGAGAGCTTGTCACTTCC
TTCTTAACGCAGTGGTTCCTATCTCAGGGATACTGGACTTGACGACACAGATGAACAAT
TAAAGTGGAAACCGCTTCCCTTTTCTCCTCTGTGGCAGTTACAATTTTGACTTCAGGCCT
GAGAAAAACTTCAGGTTTCGAAACATGACATCTCTTCCTTTTCCAGATCCCATGCTTTGA
AAAATATTTATAGACAGTTCCAGGTCTCAGCTTCCTGTCCTCTAGTTCTGCTGTTCGGGC
ATAAAATCTTTATCTCCAGTTCATATAATCTTGAGTTTTAGATAKACACACATGCGTAACA
GCTGACAGTTTTTCACAAGTACACCCACCTGTAAATACTGTAKCCTCAGTTTAGAAAATT
AGTGCATGTATGAAAATCAAGTGTTAGGAAATTTCATGGTTTCACCTATAACCTTTATTTT
AGAATTGAACTATGATTAGATTGTATCTAAACCTGAAGTATAATTATATGCAGTGCTTCTT
AAGGCTTCATAAAATAATTTTCCAACCTTTTTAAAAAAAAAAAAAAA

SEQ ID NO: 4

SAM1b peptide

MALLLAACGGLDSGFIPSVQDFDKKLTEADAYLQILIEQLKLFDDKLQNCKDDEQRKKVETLK
DTTNSMVESIKHCIVLLQIAKSTINPVDAIYQPSPLEPVISTMPSQTALPPEPAQLCKSEQRPS
SLPVGPVLATLGHHQTPTPNSTGSGNSPPSSSLTPPSHVNLSPNTVPEFSYSSSEDEFYDA
DEFHQSGSSPKRLIDSSGSASVLTHSSSGNSLKRPDTTESLNSSMSNGTSDADLFDSHDDR
DDDGEAGSVEEHKSVIMHLLSQVRLGMDLTKVVLPTFILERRSLLEMYADFFAHPDLFVSISD
QKDPRDRMVQVVKWYLSAFHAGRRGSVAKKPYNPILGEIFQCHWTLPNDTEENAELVSEG
PVPWVSKNSVTFVAEQVSHHPPISAFYAECFNKKIQFNAHIWTKSKFLGMSIGVHNIGQGCV
SCLEYDEHYILTFPNGYGRSILTVPWVELGGECNINCSKTGYSANIVFHTKPFYGGKKHRITA
EIFSPNDKKSFCSIEGEWNGIMYAKYATGENTVFVDTKKLPIIKKKVRKLEDQNEYESRSLWK
DVTFNLKIRDIDAATEAKHRLEERQRAEARERKEKEIQWETRLFHEDGECWVYDEPLLKRLG
AVKH

Novo Nordisk A/S

ABSTRACT

Two receptors of meiosis activating sterols designated SAM1a and SAM1b have been identified.

SEQUENCE LISTING

<110> NOVO NORDISK A/S

<120> SAM

<130> 6229.000-DK

<160> 4

<170> PatentIn version 3.0

<210> 1
<211> 2814
<212> DNA
<213> ARTIFICIAL SEQUENCE

<400> 1
tttgttgtca ttggcggctc ccaagatggc gtccatcgtg gaagggccgc tgagcaaatg 60

gactaacgtg atgaagggat ggcagtatcg ttggttcgtg ctggactaca atgcagggct 120

gctctcctac tacacgtcca aggacaaaat gatgagaggc tctcgaagag gatgcgttag 180

actcagagga gctgtgattg gtatagacga cgaggacgac agcaccttca caatcactgt 240

cgatcagaaa accttccact tccaggctcg agatgcagac gagcgagaga agtggatcca 300

tgccttagaa gaaactattc ttcgccatac tcttcagctt caaggtttgg attcaggatt 360

catccccagt gtccaagact ttgataagaa acttaccgag gctgacgcgt acctgcagat 420

cttgatagaa caattaaagc tttttgatga caagcttcaa aattgtaaag atgatgaaca 480

gagaaagaaa gttgaaactc tcaaagacac aacaaatagc atggtagaat caattaaaca 540

ctgcattgtg ttgctacaga ttgctaaaag tactattaat cctgtagatg caatatacca 600

gcctagtccc ttggaacctg tgatcagcac aatgccttcc cagactgcct tacctccaga 660

acccgctcag ttgtgtaagt cagagcagcg tccatcttcc ttacctgttg gacctgtgtt 720

agctaccttg ggacatcatc agactccaac accaaatagt acaggcagtg ggaactcacc 780

acctagcagc agtctgactc ctcccagcca tgtcaacttg tctccaaata cagtcccaga 840

gttctcttac tctagcagtg aagatgagtt ctatgatgct gatgaattcc atcaaagtgg 900

ctcgtcccca aagcgcttaa tagattcttc tggatctgcc tcagtcttga cacacagcag 960

ctccggaaat agcttaaaac gcccagatac cacagagtct ctgaattcct ccatgtccaa 1020

tggcacaagc gatgctgatc tttttgactc acatgacgac agagatgatg atggggaggc 1080

tgggtcagtg gaggagcaca agagcgttat catgcacctc ttatcacaag tcaggctggg 1140

gatggacctc acaaaggtag ttcttccaac gtttattctc gagagaagat ctctgttaga 1200

aatgtatgca gacttttttcg cacatccaga cctgttcgtg agcattagtg atcagaagga 1260

tcccagggat cgaatggttc aggttgtgaa atggtacctc tcggccttcc atgcaggaag 1320

gagaggatcg gtggccaaaa agccgtacaa tcctattttg ggtgagatct ttcagtgtca 1380

ctggacgttg ccgaatgata ctgaagagaa cgcagagctc gtttcagaag ggccggttcc 1440

ctgggtttct aagaacagtg taacatttgt ggctgagcaa gtttcccacc atccgcccat 1500

ttcagccttt tatgctgagt gttttaacaa gaagatacaa ttcaatgctc atatctggac 1560

taaatcaaaa ttccttggga tgtcaattgg ggtacacaac ataggtcagg gctgtgtctc 1620

gtgtctggag tacgatgagc actacatcct cacgttcccc aatggctatg gaaggtctat 1680

cctgacagtg ccctgggtgg aattgggagg agaatgcaat atcaactgct ccaaaacggg 1740

ttacagcgca aacatcgtct tccacactaa gcctttctat gggggcaaga agcacagaat 1800

tactgcagag atttttttctc cgaatgacaa gaaatccttc tgctcaattg aaggggaatg 1860

gaatggtatc atgtatgcaa aatacgcaac aggggaaaac actgtctttg tagacaccaa 1920

gaaattgcct ataatcaaga aaaaggtgag gaagttggaa gatcagaatg agtatgagtc 1980

ccgcagcctt tggaaggatg tcactttcaa tttaaaaatc agagacattg atgcagcaac 2040

ggaagcaaag cacagacttg aagaaagaca aagagcagaa gcccgagaaa ggaaggagaa 2100

ggaaattcag tgggagacga ggctctttca cgaggatggc gaatgctggg tttacgatga 2160

gcctttactg aagcgtcttg gtgctgtgaa gcattaggcc gacaaccgag tccacacctg 2220

gtgaccaggg cagtaggcgt aattaagcaa caatcgatct tccttcagga gagcttgtca 2280

cttccttctt aacgcagtgg ttcctatctc agggatactg gacttgacga cacagatgaa 2340

caattaaagt ggaaaccgct tccctttttct cctctgtggc agttacaatt ttgacttcag 2400

gcctgagaaa aacttcaggt ttcgaaacat gacatctctt ccttttccag atcccatgct 2460

ttgaaaaata tttatagaca gttccaggtc tcagcttcct gtcctctagt tctgctgttc 2520

gggcataaaa tctttatctc cagttcatat aatcttgagt tttagataka cacacatgcg 2580

taacagctga cagtttttca caagtacacc cacctgtaaa tactgtakcc tcagtttaga 2640

aaattagtgc atgtatgaaa atcaagtgtt aggaaatttc atggtttcac ctataacctt 2700

tattttagaa ttgaactatg attagattgt atctaaacct gaagtataat tatatgcagt 2760

gcttcttaag gcttcataaa ataattttcc aacctttta aaaaaaaaaa aaaa 2814

5

<210> 2
<211> 723
<212> PRT
<213> ARTIFICIAL SEQUENCE

<400> 2

Met Ala Ser Ile Val Glu Gly Pro Leu Ser Lys Trp Thr Asn Val Met
1 5 10 15

Lys Gly Trp Gln Tyr Arg Trp Phe Val Leu Asp Tyr Asn Ala Gly Leu
20 25 30

Leu Ser Tyr Tyr Thr Ser Lys Asp Lys Met Met Arg Gly Ser Arg Arg
35 40 45

Gly Cys Val Arg Leu Arg Gly Ala Val Ile Gly Ile Asp Asp Glu Asp
50 55 60

Asp Ser Thr Phe Thr Ile Thr Val Asp Gln Lys Thr Phe His Phe Gln
65 70 75 80

Ala Arg Asp Ala Asp Glu Arg Glu Lys Trp Ile His Ala Leu Glu Glu
85 90 95

Thr Ile Leu Arg His Thr Leu Gln Leu Gln Gly Leu Asp Ser Gly Phe
100 105 110

Ile Pro Ser Val Gln Asp Phe Asp Lys Lys Leu Thr Glu Ala Asp Ala
115 120 125

Tyr Leu Gln Ile Leu Ile Glu Gln Leu Lys Leu Phe Asp Asp Lys Leu
130 135 140

Gln Asn Cys Lys Asp Asp Glu Gln Arg Lys Lys Val Glu Thr Leu Lys
145 150 155 160

Asp Thr Thr Asn Ser Met Val Glu Ser Ile Lys His Cys Ile Val Leu
165 170 175

Leu Gln Ile Ala Lys Ser Thr Ile Asn Pro Val Asp Ala Ile Tyr Gln
180 185 190

Pro Ser Pro Leu Glu Pro Val Ile Ser Thr Met Pro Ser Gln Thr Ala
195 200 205

Leu Pro Pro Glu Pro Ala Gln Leu Cys Lys Ser Glu Gln Arg Pro Ser
210 215 220

Ser Leu Pro Val Gly Pro Val Leu Ala Thr Leu Gly His His Gln Thr
225 230 235 240

Pro Thr Pro Asn Ser Thr Gly Ser Gly Asn Ser Pro Pro Ser Ser Ser

```
                    245                 250                 255

Leu Thr Pro Pro Ser His Val Asn Leu Ser Pro Asn Thr Val Pro Glu
            260                 265                 270

Phe Ser Tyr Ser Ser Ser Glu Asp Glu Phe Tyr Asp Ala Asp Glu Phe
        275                 280                 285

His Gln Ser Gly Ser Ser Pro Lys Arg Leu Ile Asp Ser Ser Gly Ser
    290                 295                 300

Ala Ser Val Leu Thr His Ser Ser Ser Gly Asn Ser Leu Lys Arg Pro
305                 310                 315                 320

Asp Thr Thr Glu Ser Leu Asn Ser Ser Met Ser Asn Gly Thr Ser Asp
                325                 330                 335

Ala Asp Leu Phe Asp Ser His Asp Asp Arg Asp Asp Asp Gly Glu Ala
            340                 345                 350

Gly Ser Val Glu Glu His Lys Ser Val Ile Met His Leu Leu Ser Gln
        355                 360                 365

Val Arg Leu Gly Met Asp Leu Thr Lys Val Val Leu Pro Thr Phe Ile
    370                 375                 380

Leu Glu Arg Arg Ser Leu Leu Glu Met Tyr Ala Asp Phe Phe Ala His
385                 390                 395                 400

Pro Asp Leu Phe Val Ser Ile Ser Asp Gln Lys Asp Pro Arg Asp Arg
                405                 410                 415

Met Val Gln Val Val Lys Trp Tyr Leu Ser Ala Phe His Ala Gly Arg
            420                 425                 430

Arg Gly Ser Val Ala Lys Lys Pro Tyr Asn Pro Ile Leu Gly Glu Ile
        435                 440                 445

Phe Gln Cys His Trp Thr Leu Pro Asn Asp Thr Glu Glu Asn Ala Glu
    450                 455                 460

Leu Val Ser Glu Gly Pro Val Pro Trp Val Ser Lys Asn Ser Val Thr
465                 470                 475                 480

Phe Val Ala Glu Gln Val Ser His His Pro Pro Ile Ser Ala Phe Tyr
                485                 490                 495

Ala Glu Cys Phe Asn Lys Lys Ile Gln Phe Asn Ala His Ile Trp Thr
            500                 505                 510

Lys Ser Lys Phe Leu Gly Met Ser Ile Gly Val His Asn Ile Gly Gln
        515                 520                 525

Gly Cys Val Ser Cys Leu Glu Tyr Asp Glu His Tyr Ile Leu Thr Phe
```

```
    530                 535                 540

Pro Asn Gly Tyr Gly Arg Ser Ile Leu Thr Val Pro Trp Val Glu Leu
545                 550                 555                 560

Gly Gly Glu Cys Asn Ile Asn Cys Ser Lys Thr Gly Tyr Ser Ala Asn
                565                 570                 575

Ile Val Phe His Thr Lys Pro Phe Tyr Gly Gly Lys Lys His Arg Ile
            580                 585                 590

Thr Ala Glu Ile Phe Ser Pro Asn Asp Lys Lys Ser Phe Cys Ser Ile
        595                 600                 605

Glu Gly Glu Trp Asn Gly Ile Met Tyr Ala Lys Tyr Ala Thr Gly Glu
    610                 615                 620

Asn Thr Val Phe Val Asp Thr Lys Lys Leu Pro Ile Ile Lys Lys Lys
625                 630                 635                 640

Val Arg Lys Leu Glu Asp Gln Asn Glu Tyr Glu Ser Arg Ser Leu Trp
                645                 650                 655

Lys Asp Val Thr Phe Asn Leu Lys Ile Arg Asp Ile Asp Ala Ala Thr
            660                 665                 670

Glu Ala Lys His Arg Leu Glu Glu Arg Gln Arg Ala Glu Ala Arg Glu
        675                 680                 685

Arg Lys Glu Lys Glu Ile Gln Trp Glu Thr Arg Leu Phe His Glu Asp
    690                 695                 700

Gly Glu Cys Trp Val Tyr Asp Glu Pro Leu Leu Lys Arg Leu Gly Ala
705                 710                 715                 720

Val Lys His
```

<210> 3
<211> 2534
<212> DNA
<213> ARTIFICAL SEQUENCE

<400> 3

gggcccctcc cctgagacgg cggccggccg ccgcgaatgg ctctcctgct ggccgcttgc 60

ggaggtttgg attcaggatt catccccagt gtccaagact ttgataagaa acttaccgag 120

gctgacgcgt acctgcagat cttgatagaa caattaaagc tttttgatga caagcttcaa 180

aattgtaaag atgatgaaca gagaaagaaa gttgaaactc tcaaagacac aacaaatagc 240

atggtagaat caattaaaca ctgcattgtg ttgctacaga ttgctaaaag tactattaat 300

cctgtagatg caatatacca gcctagtccc ttggaacctg tgatcagcac aatgccttcc 360

cagactgcct tacctccaga acccgctcag ttgtgtaagt cagagcagcg tccatcttcc 420

ttacctgttg gacctgtgtt agctaccttg ggacatcatc agactccaac accaaatagt 480

acaggcagtg ggaactcacc acctagcagc agtctgactc ctcccagcca tgtcaacttg 540

tctccaaata cagtcccaga gttctcttac tctagcagtg aagatgagtt ctatgatgct 600

gatgaattcc atcaaagtgg ctcgtcccca aagcgcttaa tagattcttc tggatctgcc 660

tcagtcttga cacacagcag ctccggaaat agcttaaaac gcccagatac cacagagtct 720

ctgaattcct ccatgtccaa tggcacaagc gatgctgatc tttttgactc acatgacgac 780

agagatgatg atggggaggc tgggtcagtg gaggagcaca agagcgttat catgcacctc 840

ttatcacaag tcaggctggg gatggacctc acaaaggtag ttcttccaac gtttattctc 900

gagagaagat ctctgttaga aatgtatgca gacttttttcg cacatccaga cctgttcgtg 960

agcattagtg atcagaagga tcccagggat cgaatggttc aggttgtgaa atggtacctc 1020

tcggccttcc atgcaggaag gagaggatcg gtggccaaaa agccgtacaa tcctattttg 1080

ggtgagatct ttcagtgtca ctggacgttg ccgaatgata ctgaagagaa cgcagagctc 1140

gtttcagaag ggccggttcc ctgggtttct aagaacagtg taacatttgt ggctgagcaa 1200

gtttcccacc atccgcccat ttcagccttt tatgctgagt gttttaacaa gaagatacaa 1260

ttcaatgctc atatctggac taaatcaaaa ttccttggga tgtcaattgg ggtacacaac 1320

ataggtcagg gctgtgtctc gtgtctggag tacgatgagc actacatcct cacgttcccc 1380

aatggctatg gaaggtctat cctgacagtg ccctgggtgg aattgggagg agaatgcaat 1440

atcaactgct ccaaaacggg ttacagcgca aacatcgtct tccacactaa gcctttctat 1500

gggggcaaga agcacagaat tactgcagag atttttttctc cgaatgacaa gaaatccttc 1560

tgctcaattg aaggggaatg gaatggtatc atgtatgcaa aatacgcaac aggggaaaac 1620

actgtctttg tagacaccaa gaaattgcct ataatcaaga aaaaggtgag gaagttggaa 1680

gatcagaatg agtatgagtc ccgcagcctt tggaaggatg tcactttcaa tttaaaaatc 1740

agagacattg atgcagcaac ggaagcaaag cacagacttg aagaaagaca aagagcagaa 1800

gcccgagaaa ggaaggagaa ggaaattcag tgggagacga ggctctttca cgaggatggc 1860

gaatgctggg tttacgatga gcctttactg aagcgtcttg gtgctgtgaa gcattaggcc 1920

gacaaccgag tccacacctg gtgaccaggg cagtaggcgt aattaagcaa caatcgatct 1980

tccttcagga gagcttgtca cttccttctt aacgcagtgg ttcctatctc agggatactg 2040

gacttgacga cacagatgaa caattaaagt ggaaaccgct tccctttttct cctctgtggc 2100

agttacaatt ttgacttcag gcctgagaaa aacttcaggt ttcgaaacat gacatctctt 2160

ccttttccag atcccatgct ttgaaaaata tttatagaca gttccaggtc tcagcttcct 2220

gtcctctagt tctgctgttc gggcataaaa tctttatctc cagttcatat aatcttgagt 2280

tttagataka cacacatgcg taacagctga cagtttttca caagtacacc cacctgtaaa 2340

tactgtakcc tcagtttaga aaattagtgc atgtatgaaa atcaagtgtt aggaaatttc 2400

atggtttcac ctataacctt tattttagaa ttgaactatg attagattgt atctaaacct 2460

gaagtataat tatatgcagt gcttcttaag gcttcataaa ataattttcc aacctttta 2520

aaaaaaaaaa aaaa 2534

<210> 4
<211> 626
<212> PRT
<213> ARTIFICIAL SEQUENCE

<400> 4

Met Ala Leu Leu Leu Ala Ala Cys Gly Gly Leu Asp Ser Gly Phe Ile
1 5 10 15

Pro Ser Val Gln Asp Phe Asp Lys Lys Leu Thr Glu Ala Asp Ala Tyr
20 25 30

Leu Gln Ile Leu Ile Glu Gln Leu Lys Leu Phe Asp Asp Lys Leu Gln
35 40 45

Asn Cys Lys Asp Asp Glu Gln Arg Lys Lys Val Glu Thr Leu Lys Asp
50 55 60

Thr Thr Asn Ser Met Val Glu Ser Ile Lys His Cys Ile Val Leu Leu
65 70 75 80

Gln Ile Ala Lys Ser Thr Ile Asn Pro Val Asp Ala Ile Tyr Gln Pro
85 90 95

Ser Pro Leu Glu Pro Val Ile Ser Thr Met Pro Ser Gln Thr Ala Leu
100 105 110

Pro Pro Glu Pro Ala Gln Leu Cys Lys Ser Glu Gln Arg Pro Ser Ser
115 120 125

Leu Pro Val Gly Pro Val Leu Ala Thr Leu Gly His His Gln Thr Pro
130 135 140

Thr Pro Asn Ser Thr Gly Ser Gly Asn Ser Pro Pro Ser Ser Ser Leu
145 150 155 160

Thr Pro Pro Ser His Val Asn Leu Ser Pro Asn Thr Val Pro Glu Phe
165 170 175

Ser Tyr Ser Ser Ser Glu Asp Glu Phe Tyr Asp Ala Asp Glu Phe His
180 185 190

Gln Ser Gly Ser Ser Pro Lys Arg Leu Ile Asp Ser Ser Gly Ser Ala
195 200 205

Ser Val Leu Thr His Ser Ser Ser Gly Asn Ser Leu Lys Arg Pro Asp
210 215 220

Thr Thr Glu Ser Leu Asn Ser Ser Met Ser Asn Gly Thr Ser Asp Ala
225 230 235 240

Asp Leu Phe Asp Ser His Asp Asp Arg Asp Asp Asp Gly Glu Ala Gly
245 250 255

```
Ser Val Glu Glu His Lys Ser Val Ile Met His Leu Leu Ser Gln Val
            260                 265                 270

Arg Leu Gly Met Asp Leu Thr Lys Val Val Leu Pro Thr Phe Ile Leu
        275                 280                 285

Glu Arg Arg Ser Leu Leu Glu Met Tyr Ala Asp Phe Phe Ala His Pro
    290                 295                 300

Asp Leu Phe Val Ser Ile Ser Asp Gln Lys Asp Pro Arg Asp Arg Met
305                 310                 315                 320

Val Gln Val Val Lys Trp Tyr Leu Ser Ala Phe His Ala Gly Arg Arg
                325                 330                 335

Gly Ser Val Ala Lys Lys Pro Tyr Asn Pro Ile Leu Gly Glu Ile Phe
            340                 345                 350

Gln Cys His Trp Thr Leu Pro Asn Asp Thr Glu Glu Asn Ala Glu Leu
        355                 360                 365

Val Ser Glu Gly Pro Val Pro Trp Val Ser Lys Asn Ser Val Thr Phe
    370                 375                 380

Val Ala Glu Gln Val Ser His His Pro Pro Ile Ser Ala Phe Tyr Ala
385                 390                 395                 400

Glu Cys Phe Asn Lys Lys Ile Gln Phe Asn Ala His Ile Trp Thr Lys
                405                 410                 415

Ser Lys Phe Leu Gly Met Ser Ile Gly Val His Asn Ile Gly Gln Gly
            420                 425                 430

Cys Val Ser Cys Leu Glu Tyr Asp Glu His Tyr Ile Leu Thr Phe Pro
        435                 440                 445

Asn Gly Tyr Gly Arg Ser Ile Leu Thr Val Pro Trp Val Glu Leu Gly
    450                 455                 460

Gly Glu Cys Asn Ile Asn Cys Ser Lys Thr Gly Tyr Ser Ala Asn Ile
465                 470                 475                 480

Val Phe His Thr Lys Pro Phe Tyr Gly Gly Lys Lys His Arg Ile Thr
                485                 490                 495

Ala Glu Ile Phe Ser Pro Asn Asp Lys Lys Ser Phe Cys Ser Ile Glu
            500                 505                 510

Gly Glu Trp Asn Gly Ile Met Tyr Ala Lys Tyr Ala Thr Gly Glu Asn
        515                 520                 525

Thr Val Phe Val Asp Thr Lys Lys Leu Pro Ile Ile Lys Lys Lys Val
    530                 535                 540
```

```
Arg Lys Leu Glu Asp Gln Asn Glu Tyr Glu Ser Arg Ser Leu Trp Lys
545                 550                 555                 560

Asp Val Thr Phe Asn Leu Lys Ile Arg Asp Ile Asp Ala Ala Thr Glu
                565                 570                 575

Ala Lys His Arg Leu Glu Glu Arg Gln Arg Ala Glu Ala Arg Glu Arg
            580                 585                 590

Lys Glu Lys Glu Ile Gln Trp Glu Thr Arg Leu Phe His Glu Asp Gly
        595                 600                 605

Glu Cys Trp Val Tyr Asp Glu Pro Leu Leu Lys Arg Leu Gly Ala Val
    610                 615                 620

Lys His
625
```

Novo Nordisk A/S